

Compressed Sensing in astronomy and remote sensing: a data fusion perspective

J. Bobin^a and J.-L. Starck^b,

^a Department of Applied & Computational Mathematics
California Institute of Technology, Pasadena, California 91125, USA

^bCEA, IRFU, SEDI-SAP, Lab. AIM (UMR 7158) ,
CEA/DSM-CNRS-Universite Paris Diderot, Centre de Saclay,
F-91191 Gif- Sur-Yvette, France.

Keywords: Compressed sensing, astronomy, remote sensing, data fusion, wavelets, data fusion

ABSTRACT

Recent advances in signal processing have focused on the use of sparse representations in various applications. A new field of interest based on sparsity has recently emerged : *compressed sensing*. This theory is a new sampling framework that provides an alternative to the well-known Shannon sampling theory. In this paper we investigate how *compressed sensing* (CS) can provide new insights into astronomical data compression. In a previous study¹ we gave new insights into the use of Compressed Sensing (CS) in the scope of astronomical data analysis. More specifically, we showed how CS is flexible enough to account for particular observational strategies such as raster scans. This kind of CS data fusion concept led to an elegant and effective way to solve the problem ESA is faced with, for the transmission to the earth of the data collected by PACS, one of the instruments onboard the Herschel spacecraft which will be launched in late 2008/early 2009.

In this paper, we extend this work by showing how CS can be effectively used to jointly decode multiple observations at the level of map making. This allows us to directly estimate large areas of the sky from one or several raster scans. Beyond the particular but important Herschel example, we strongly believe that CS can be applied to a wider range of applications such as in earth science and remote sensing where dealing with multiple redundant observations is common place. Simple but illustrative examples are given that show the effectiveness of CS when decoding is made from multiple redundant observations.

Keywords: compressed sensing, sparsity, remote sensing, wavelets, astronomy.

Introduction

From year to year, the quantity of astronomical data increases at an ever growing rate. In part this is due to very large digital sky surveys in the optical and near infrared, which in turn has been made possible by the development of digital imaging arrays such as CCDs (charge-coupled devices). The size of digital arrays is continually growing, pushed by the demands of astronomical research for ever larger quantities of data in ever shorter time periods. As a result, the astronomical community is also confronted with a rather desperate need for data compression techniques. Several techniques have in fact been used, or even developed, for astronomical data compression^{2,3}. For some projects, we need to achieve huge compression ratios, which cannot be obtained by current methods without introducing unacceptable distortions. Furthermore, for most astronomical data compression problems, three main properties must be under control : resolution (point spread function), sensitivity (ability to detect low level signals) and photometry.

Further author information: Send correspondence to bobin@acm.caltech.edu.

This work was partially supported by the French National Agency for Research (ANR -08-EMER-009-01).

The Herschel satellite*, which will have been launched in May 2009, is faced with a similar problem. Indeed the photometer data need to be compressed by a factor of approximately 8 to be transferred. As the CPU load has to be extremely small, conventional compression methods cannot be used.

Recently, an alternative sampling theory has emerged which shows that signals can be recovered from far fewer samples (measurements) than what the Nyquist/Shannon sampling theory states. This new theory coined *compressed sensing* or (compressive sensing) introduced in the seminal papers^{4,5} relies on the compressibility of signals or more precisely on the property for some signals to be sparsely represented. From the compressed sensing viewpoint, sparse signals could be acquired “economically” (from a few samples) *without loss of information*. It introduces new conceptions in data acquisition and sampling. Beyond the scope of astronomical data analysis, merging highly redundant data or multiple observations is common place in geoscience and remote sensing. In that case, Compressed sensing turns out to be to cope with morphologically complex and high dimension data compression.

Scope of the paper : We review in Section 1 the main leading concepts of Compressed sensing. In Section 2 some CS results from a data fusion perspective. More particularly, we extend the previous study¹ in Section 2.1 by showing how jointly decoding raster scans data is made effective in the CS framework. In other words, we show that combining astronomical data map making and CS decoding yields efficient recovery results. Beyond the scope of astronomical data analysis, we present a simple but illustrative comparison of CS with a standard wavelet-based compression scheme from a data fusion perspective. In Section 2.2, numerical experiments are given which show that CS outperform a wavelet-based compression techniques when multiple observations of the same datum are involved.

1. COMPRESSED SENSING, AN OVERVIEW

In this section, we give a brief and non-exhaustive review of compressed sensing and show how this new sampling theory will probably lead to a “revolution” in signal processing and communication theory. For more exhaustive tutorials in this field, we refer the reader to the review papers.^{6,7} In this paper, we will assume that the signal x belongs to \mathbb{R}^t (written as a column vector with t entries or samples). We will also assume that x is **compressible**.

1.1. The essence of compressed sensing

Compressibility The content of most astronomical images is often well structured : diffuse gas clouds, point sources ... to name only a few. Recent advances in harmonic analysis have provided tools that efficiently represent such structures (wavelets, curvelets, ... to name a few). In this context, efficient representations mean sparse representations. Let us consider a signal x of size t . Let Φ be an orthonormal basis (*e.g.* classically, an orthogonal wavelet basis for astronomical data processing) and let us consider the projection of x in Φ :

$$\alpha = \Phi^T x \quad (1)$$

The signal x is said to be **sparse** in Φ if most entries of the so-called coefficient vector α are zero or close to zero and thus only a few have significant amplitudes. In other words, the signal x can be efficiently approximated (with low ℓ_2 approximation error or distortion) from only a few significant coefficients. Then such sparse signal is said to be *compressible*. Note that, in the last decade, sparsity has emerged as one of the leading concepts in a wide range of signal processing applications. More formally, we will distinguish two categories of compressible signals :

- **Strict sparsity** : only $K \ll t$ entries of α are different from zero. x is said to be K -sparse in Φ .
- **Wide sense compressibility** : A more realistic definition of compressibility consists in describing how the entries of α behave. Let us consider that, $\alpha[1] \geq \dots \geq \alpha[t]$. Then x is said to be compressible in Φ , if there exists $p \in]0, 1]$ such that : $|\alpha[i]| \leq C i^{-1/p}$. Here p defines a kind of sparsity or compressibility degree. Real-world data are more akin to be wide sense compressible.

*See <http://www.esa.int/science/herschel>

Collecting incoherent measurements Assuming that x is compressible (*i.e.* x has a sparse representation in a particular basis Φ), x can be efficiently approximated (with low distortion) from only a few entries of $\alpha = \Phi^T x$. In the Compressed Sensing framework, the signal x is not acquired directly; it is *observed* or *measured* from $M < t$ *measurements* $\{y_k\}_{k=1,\dots,M}$. These measurements are obtained by projecting the signal x on a set of so-called measurement vectors $\{\theta_k\}_{k=1,\dots,M}$ as follows :

$$y_k = \langle x, \theta_k \rangle \quad (2)$$

Each sample y_k is then the scalar product of x with a specific *measurement* vector θ_k . The gist of compressed sensing relies on two major concepts : i) the data to be compressed are indeed compressible; more precisely the data x have a “structured” content so that they can be sparsely represented in some basis Φ ; ii) the measurement vectors $\{\theta_k\}_{k=1,\dots,M}$ are non adaptive (they should not depend on x) and incoherent with the basis in which x is assumed to be sparse.

As stated in,⁶ two main categories of measurement ensembles can be used for CS coding :

- **Random measurements** : Θ is not explicitly used; the measurements $\{y_k\}_{k=1,\dots,M}$ are random linear combinations of the entries of x . Fourier, Binary or Gaussian measurements are widely used. In the CS framework, incoherent measurements can be obtained by using random ensembles (see^{4,8} and references therein). Randomness is likely to provide *incoherent* projections.
- **Incoherent measurements** : In that case, Θ is a deterministic basis which is assumed to be *incoherent* with Φ . More quantitatively, incoherence between Θ and Φ is measured by their mutual coherence : $\mu = \max_{i,k} |\langle \phi_i, \theta_k \rangle|$. The lower μ is, the more incoherent Θ and Φ are. In practical situations, typical astronomical data are compressible in a wavelet basis Φ ; a good choice for Θ is the noiselet basis.⁹

In this paper, measurement vectors are designed by selecting at random a set (indexed by Λ) of vectors from a deterministic orthonormal basis Θ as suggested in¹⁰ : $y = \Theta_\Lambda x$.

An empirical interpretation : In the CS framework, the signal to be transferred is $y = \Theta\Phi\alpha$. Let recall that the backbone of CS is twofold :

- **the data are compressible** : it means that only a few entries of α have a significant amplitude; x is then almost entirely determined from only a few entries α .
- **the measurements are incoherent** : the measurement matrix Θ and Φ are incoherent. From a empirical point of view, the incoherence of Θ and Φ means that the information carried by a few entries of α is spread all over the M entries of $y = \Theta\Phi\alpha$. Each sample y_k is likely to contain a piece of information of each significant entry of x . As $M < t$, the ratio $\rho = M/t$ is equivalent to a compression ratio.

1.2. Signal recovery

Exact solutions The previous paragraph emphasized on the way the compression step should be devised. The decompression step amounts to recovering the original signal x out of the compressed signal $y = \Theta_\Lambda x$. Furthermore, x is known *a priori* to be compressible in Φ . Then the recovery problem boils down to emphasizing on the sparsity of the vector $\alpha = x\Phi$. As proposed in,^{4,5} the decompression or decoding step is equivalent to solving the following optimization problem :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } y = \Theta_\Lambda \Phi \alpha \quad (3)$$

Several strong recovery results in the particular CS framework have been proved based on specific assumptions with random measurement ensembles (see^{4,11}). For instance, in the extreme strict sparsity

case where only K entries of α are non-zero, conditions are given⁴ showing that the problem in Equation (3) provides the *exact* solution α . Nevertheless, the data are often corrupted by noise. A more realistic compression model would be the following :

$$y = \Theta_{\Lambda}(x + n) \quad (4)$$

where n is a white Gaussian noise with variance σ_n^2 . As the measurement matrix Θ_{Λ} is a sub-matrix of the orthonormal matrix Θ , the projected noise $n_{\Lambda} = \Theta_{\Lambda}n$ is still white and Gaussian with the same variance σ_n^2 . The projected data are then recast as follows : $y = \Theta_{\Lambda}x + n_{\Lambda}$. The recovery step then boils down to solving the next optimization problem :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } \|y - \Theta_{\Lambda}\Phi\alpha\|_{\ell_2} \leq \epsilon \quad (5)$$

where ϵ is an upper bound of $\|n\|_{\ell_2}$. Defining $\epsilon = \sqrt{M + 2\sqrt{2M}}\sigma_n$ provides a reasonable upper bound on the noise ℓ_2 norm, with overwhelming probability. Conditions¹² are given that guarantee some optimality results from the problem in Equation (5).

The convex program (second-order cone program) in Equation (5) then provides an efficient and robust mechanism to provide an approximate to the signal x .

2. A DATA FUSION PERSPECTIVE IN ASTRONOMY AND REMOTE SENSING

In the wide field of remote sensing ranging from astronomical to earth survey, the observation of redundant data is common (just think of overcomplete raster scan data). Designing a compression scheme such that i) coding is computationally inexpensive and easy to implement on board a satellite and ii) decoding is able to take into account the redundancy of the data, is challenging. Off-the-shelf compression techniques such as JPEG or JPEG 2000 are very efficient but their computation cost makes their use on board almost impossible. In this section, we want to stress the ability of compressed sensing to fulfill the above requirements. More precisely, we want to convince the reader that CS can be the appropriate tool for a low onboard computational cost while outperforming state-of-the-art image compression techniques when the data are redundant.

More precisely, let us think of a raster scan of a given area of the sky X . The raster scan delivers a sequence of L highly overlapping images $\{x_i\}_{i=1,\dots,L}$ each of which being a small area Ω_i of X . Let \mathcal{P}_{Ω_i} the orthogonal projector onto Ω_i^{\dagger} ; then each signal x_i is no more than :

$$x_i = \mathcal{P}_{\Omega_i}(X) \quad (6)$$

Let recall that, in the CS framework, the measurements y are obtained by computing the scalar product of a set of measurement vectors and the signal to be compressed. This process being linear, if one has a set of measurement vectors $\{y_i\}_{i=1,\dots,L}$ that have been computed with different measurement matrices $\{\Theta_{\Lambda_i}\}_{i=1,\dots,L}$, then decoding the signal x from the set $\{y_i\}_{i=1,\dots,L}$ should, intuitively, be helpful to reconstruct the sky area X of size $T \times T$.

For simplicity, let each measurement matrix Θ_i be such that :

$$\Theta_{\Lambda_i} = \Lambda_i \Theta \quad (7)$$

where Θ is an orthogonal matrix and Λ_i is made of a M rows of the identity matrix of size t . Basically, each measurement matrix is such that its Gram matrix forms an orthogonal projector and is made of a few rows of the matrix Θ . Each measurement y_i is then computed as follows :

$$y_i = \Lambda_i \Theta x_i + n_i \quad (8)$$

[†]In the following, Ω_i will denote either a subset of pixels or their indices.

where n_i stands for a noise term which we assume to be white, Gaussian and of variance σ^2 . As each image x_i is a small patch of the sky area X , then each measurement vector y_i is related to X as follows :

$$y_i = \mathbf{\Lambda}_i \mathbf{\Theta} \mathcal{P}_{\Omega_i}(X) + n_i \quad (9)$$

Assuming that the sky area X is sparse in Φ , it can then be reconstructed from the set of measurements $\{y_i\}_{i=1,\dots,L}$ as follows :

$$\min_X \|\Phi^T X\|_{\ell_1} \text{ s.t. } \sum_{i=1}^L \|y_i - \mathbf{\Lambda}_i \mathbf{\Theta} \mathcal{P}_{\Omega_i}(X)\|_{\ell_2}^2 \leq \epsilon^2 \quad (10)$$

where choosing $\epsilon = \sqrt{LM + 2\sqrt{2LM}\sigma}$ is reasonable as the noise is assumed to be white and Gaussian. Note that the full sensing process $X \rightarrow \{y_i\}_{i=1,\dots,L}$ is totally linear. Let Y be the vector obtained by vectorizing the set of measurement vectors $\{y_i\}_{i=1,\dots,L}$. Then its not hard to see that there exists a $LM \times T$ matrix \mathbf{S} such that

$$Y = \mathbf{S}X + n \quad (11)$$

Let us remark that the equivalent “sensing” matrix \mathbf{S} can be easily defined by blocks of size $M \times T$ $\{\mathbf{S}_i\}_{i=1,\dots,L}$:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_L \end{bmatrix} \quad (12)$$

Let $\mathbf{\Omega}_i$ the $t \times T$ matrix formed by extracting the submatrix indexed by Ω_i of the identity matrix of size $T \times T$. Similarly $\mathbf{\Lambda}_i$ the $M \times t$ matrix formed by extracting the submatrix indexed by Λ_i of the identity matrix of size $t \times t$; $\mathbf{\Lambda}_i$ has also been defined earlier. Then each submatrix \mathbf{S}_i is formed as follows :

$$\mathbf{S}_i = \mathbf{\Lambda}_i \mathbf{\Theta} \mathbf{\Omega}_i \quad (13)$$

The problem in Equation (10) can then be straightforwardly recast as

$$\min_X \|\Phi^T X\|_{\ell_1} \text{ s.t. } \|Y - \mathbf{S}X\|_{\ell_2}^2 \leq \epsilon^2 \quad (14)$$

which is a quadratically constrained ℓ_1 -minimization problem in its standard formulation. This can be solved by using recently introduced fast algorithms such as NESTA.¹³

Intuitively, recovering X from the L measurements should be better than recovering each signal x_i independently and merging the recovered signals to get X . To give the gist of this intuition, just consider that recovering each image x_i independently would be made from y_i only (that is $M \ll t$) measurements; recovering X directly is made from $L \times M$ measurements. When the data are highly redundant and the sky patches x_i are highly overlapping, the effective number of measurements M' can be very high thus entailing potentially good decoding performances. Notice that in a previous study¹ we presented some preliminary results in the case $x_i = X$. We showed with a simple toy example that reconstructing X from a large number of redundant data is highly profitable. In the next, we show how this “data fusion”-based reconstruction applies in the scope of Herschel data processing in Section 2.1. We strongly believe that CS could be applied to a wider of of remote sensing surveys such as in geoscience; to give support to this idea, we propose a simple comparison of CS and JPEG 2000 in Section 2.2.

2.1. Compressed sensing in astronomy, the Herschel case

2.1.1. The Herschel dilemma

Herschel is one of the cornerstone missions of the European Space Agency (ESA). This space telescope has been designed to observe in the far-infrared and sub-millimeter wavelength range. Its has been launched

in early May 2009. The shortest wavelength band, 57-210 μm , is covered by PACS (Photodetector Array Camera and Spectrometer),¹⁴ which provides low to medium resolution spectroscopy and dual-band photometry. When PACS is used as a photometer, it will simultaneously image with its two bolometer arrays, a 64×32 and a 32×16 matrix, both read out at 40 Hz. The ESA is faced with a challenging problem : conventional low-cost compression techniques cannot achieve a satisfactory compression rate. In this Section, we propose a new CS-based compression scheme for the Herschel/PACS data that yield an elegant and effective way to overcome the Herschel compression dilemma.

The Herschel space telescope is partially hampered by the narrowness of the transmission band compared to the large amount of data to be transferred. This handicap stems from the limitation of conventional compression techniques to provide adequate compression rate with low computational cost, given the high readout noise. More quantitatively, the data have to be compressed in real time by a factor of 6 with very low CPU power.

Problem statement The Herschel spacecraft is about to be launched; ESA is faced with a compression problem as the data need to be compressed by a factor of $N = 8$. Up to now, the only acceptable solution for the ESA (with respect to computational cost and quality) to overcome this need for a higher compression rate is the average of $N = 8$ consecutive images.¹⁵ Indeed, the compression code has no information about the scan speed or the scan direction and a shift-and-add averaging solution is not possible. Other compression techniques such JPEG or JPEG2000 are also not acceptable because of computation time constraints.

Herschel will observe wide sky areas thus requiring fast scanning strategies. Herschel/PACS will provide sets of consecutive 64×32 images that will be shifted with a typical shift value $\lambda = 1$ pixel in fast scanning mode. Unfortunately, the shift value λ is comparable to the FWHM (full width at half maximum) of the instrumental PSF (point spread function) is $\delta \simeq 3$ pixels. As a consequence, averaging N consecutive images will entail a catastrophic loss of spatial resolution. This can be catastrophic for some scientific programs. Furthermore, averaging is far less optimal for noise reduction as averaging shifted signals does not yield a \sqrt{N} noise variance reduction.

An effective compression scheme would have to balance between the following performance criteria :

- **Spatial resolution :** averaging fast scanned data entails a lower spatial resolution. An effective compression scheme should provide a lower resolution loss.
- **Sensitivity :** averaging will reduce noise but will also blur the data thus entailing a loss of sensitivity. Sensitivity (*i.e.* ability to detect low level details or sources) after compression/decompression must be under control.

These two points have been studied in a previous paper.¹ We have particularly showed on some toy examples that Compressed sensing could outperform ESA's solution both in terms of Spatial resolution and sensitivity. More precisely, we illustrated that CS could yield an increase of the spatial resolution up to 30% while keeping at least the same sensitivity. In the next section, we extend this study in two ways :

- We study the impact of CS on photometry; that is its effect on the estimation of the flux of the point sources.
- From a data fusion perspective, the next study consists in recovering a simulated sky area instead of simply decoding a single image from shifted versions as performed previously.¹ This is then the first step towards a coherent decoding of multiple observations at the level of map making.

2.1.2. Results

In the scope of astronomical data processing, reconstructing the full sky X is no more than performing the usual map making step. As we mentioned above, decoding the data at the level of map making should yield decoding performances that are better than simply decoding each measurement y_i independently before performing the map making step. In this section, we illustrate the effectiveness of CS in accounting for the redundancy of the data in the scope a simple toy-example.

Let X be a 64×256 image representing an idealized portion of the sky. X is made of 64 point sources convolved with the Gaussian PSF of width 3 pixels. Each point source has a random amplitude a generated uniformly in log-scale so that it ranges in an interval of 10 decades. The point sources are generated such that each one lies alone in a box of size 16×16 pixels (which will make the computation of their flux easy). An example of such data is shown in Figure 1.

The sky area X is observed via a raster scan composed of $L = 224$ images x_i of size 64×32 . Each consecutive pair x_i and x_{i+1} is only shifted by a single pixel along the horizontal axis (along which the length of X is 256 pixels). Each measurement vector y_i is obtained by selecting $1/8$ of the Hadamard[‡] coefficient of x_i . This selection is made randomly among 8 non-overlapping sets of indices $\{\Lambda_i\}_{i=1,\dots,8}$. White Gaussian noise is added with $\sigma = 10$. CS decoding is performed by solving the problem in Equation (14) with NESTA.¹³

We assume that the same sky area X is observed several times (from 1 to 25 times). We compare CS with some ideal but impossible to implement onboard Wavelet compression scheme. Let us assume that we are able to perform onboard the wavelet decomposition of each observation x_i , the somehow optimal compression scheme consists in selecting the best M terms (most significant wavelet coefficients). Let us note that we neglect the coding of their position which would slightly decrease the overall compression ratio.

In astronomy, recovering the flux of the point sources with high accuracy is crucial. We then propose to compare CS and the simple wavelet compression scheme described above on the basis of flux estimation. The results are displayed in Figure 2; the picture at the top shows the value of the reference flux (true flux) versus the estimated flux when only one raster scan is performed. Notice that, this figure shows the results we got from 3 random trials of X . We see that both methods seem to perform similarly for flux values ranging from 10^5 to 10^9 (4 decades). The picture at the bottom displays the same results obtained from 25 raster scans. Quantitative results are given in Table 3; each entry is the mean normalized flux \bar{F} . Let f_i (respectively \tilde{f}_i) be the reference flux (*resp.* estimated flux) of the i -th source, then \bar{F} is defined as follows :

$$\bar{F} = 1/p \sum_{i=1}^p |f_i - \tilde{f}_i| / f_i \quad (15)$$

where p is the number of point sources. The main advantage of this mean normalized flux is its sensitivity to the error on the low flux values.

Basically, the performances of the wavelet compression scheme does not change dramatically as it selects high amplitude coefficients that are slightly perturbed by noise but its non-linearity makes it hard to get enhanced results when the data are redundant. In other words, dim point sources are definitively thrown away and can't be recovered whatever the number of scans is. CS does not perform similarly to the wavelet compression scheme for the low flux values (as shown in Table 3). Owing to the linearity of CS coding, decoding from redundant measurements (25 raster scans) provides recovery results that

[‡]The Hadamard transform is also a candidate transform for the Herschel/PACS data compression.

outperform the wavelet compression scheme. Note also that, even when decoding is made from a single raster scan, CS provides results that are similar to wavelet compression for flux values in the range $[10^5, 10^9]$.

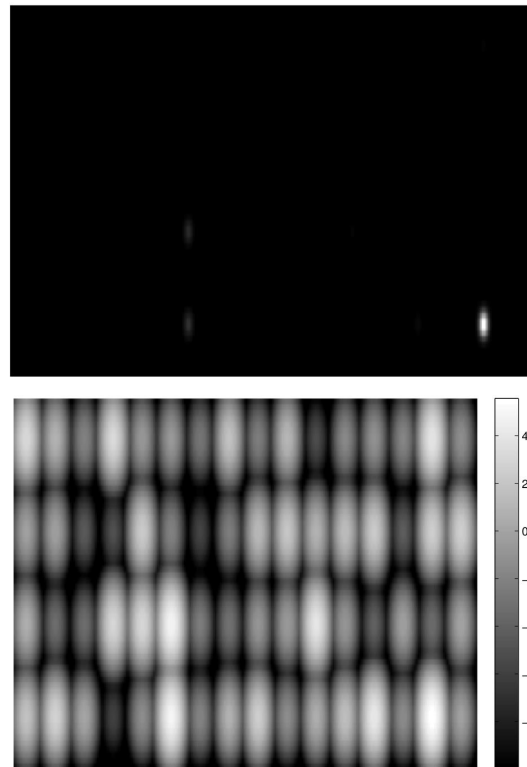


Figure 1. **Top** : simulated image X . **Bottom** : simulated image X in log scale.

# Scans	1	10	25
Wavelets	$1.19 \cdot 10^6$	$1.19 \cdot 10^6$	$1.19 \cdot 10^6$
CS	$2.46 \cdot 10^6$	$1.93 \cdot 10^6$	$0.65 \cdot 10^6$

Table 1. Photometry : Mean normalized flux error for CS and wavelet compression for varying number of raster scans.

Discussion on Herschel/PACS data processing When dealing with Herschel data, this simple experiment is far from being realistic. Nevertheless it gives more support to the use of CS to compress Herschel data. More realistic models should include the flat field estimation (described previously¹), a more sophisticated modeling of the relation between consecutive images in a raster scan : beyond a simple one-pixel shift, it should account for geometrical transformations (rotations, ...etc) which is common when performing the map making step of astronomical data.

A key point that has not been mentioned is quantization. The measurements $\{y_i\}_{i=1,\dots,L}$ which are eventually transmitted are quantized. The effect of quantization on the recovery performances should be evaluated. Quantization could also be accounted for in Equation 14 by substituting the ℓ_2 norm with a ℓ_∞ norm.

These points will be the topic of future communications.

2.2. Compressed sensing and data fusion

In the previous section, we showed that performing map-making and CS-decoding jointly yield effective compression results compared to more conventional but hardly applicable compression techniques. More particularly, the map making process can be part of the CS decoding process. Believing that this concept can be extended to other fields such as geoscience, we would like to give in this section a simple experiment on realistic data commonly observed in earth science. Let's consider the top-left picture in Figure 3, it shows an real-world aerial image of size 256×256 pixels. Contrary to most astronomical data, this kind of image is composed of very diverse morphological features : texture parts, contours of varying length and width, ...etc.

Furthermore, the imaging sensors tends to have growing numbers of pixels thus entailing a need in effective compression techniques. Due to onboard computational limitations, the use of state-of-the-art image compression schemes such as JPEG or JPEG 2000 is almost impossible.

Next, we propose comparing Compressed Sensing with a best M -term approximation in a wavelet orthogonal basis; the latter is somehow the first test towards the JPEG 2000 compression scheme. The top-right picture of Figure 3 displays the best M -term approximation of the aerial image we mentioned above for a "compression" ratio $\rho = t/M = 5$. Let's consider now that we observe x , the same earth area, L times (ideally, up to noise and some geometrical deformation). The wavelet-based compression technique (here a shortcut for the best M -term approximation in an orthogonal basis) always picks up non-linearly the same most significant M terms. Recovering x from multiple observations does not help. Owing to the linearity of the CS coding, x can be estimated by jointly decoding the L observations. This can be done by solving the problem in Equation 14. In the following experiment, each measurement vector y_i is computed by randomly selecting $M = t/\rho$ Hadamard coefficients of x (ρ is the "compression ratio"). We assume the image x is sparse in an undecimated wavelet transform Φ . Coding and decoding are performed for 31 different numbers of observations $L \in \{1, \dots, 100\}$ and 31 different values of the compression ratio $\rho \in \{1.03, \dots, 100\}$. CS decoding is performed by solving the problem in Equation (14) with NESTA.¹³

The two pictures at the bottom of Figure 3 depict the CS recovery results from 2 and 10 observations in the case $\rho = 5$. At first sight, the CS recovery from 2 observations is outperformed by the wavelet-based compression technique; when the decoding is made from 10 observations, the results look visually better than the wavelet-based compression results. More quantitatively, Figure 4 displays the evolution of the SNR recovery when the number of observations L ranges from 1 to 100 in the case $\rho = 5$. We can observe that when $L > 7$, CS outperforms Wavelet compression. Note that when L is high enough (here $L > 60$), the SNR does not improve. In that case, the set of measurements $\{y_i\}_{i=1,\dots,L}$ is likely to contain exactly all the Hadamard coefficients of x ; adding more observations is then useless. Note that this is a noiseless experiment, if additive independent (from one observation to another) noise were added, noise would be "averaged" when merging different observations. In that case, decoding from an increasing number of observations would be even better.

Figure 5 shows the number of observations N_{CS} needed for CS to outperform the wavelet-based compression scheme when the compression ratio ρ ranges from 1.03 to 100. Clearly we see that for reasonable values of the compression ratio $\rho < 20$, the number of observations need for CS to outperform the Wavelet compression is also reasonable (typically lower than 15).

3. CONCLUSION

In a previous paper¹ we presented some new insights into the application of Compressed Sensing in astronomy and more particularly astronomical data compression. In this paper, we show how CS can revolutionize astronomical data compression and processing : we particularly put emphasis on the ability of CS to account for multiple observations such as those provided by raster scans. In this case, which is common place in astronomy, redundant data can be decoded jointly within the map making process. This makes CS more flexible and effective than off-the-shelf standard techniques such as wavelet-based compression. Beyond the scope of astronomy, we present some simple but illustrative examples that could arise in earth science where merging redundant data is far from being rare. Numerical experiments are given which show the effectiveness of CS to jointly handle multiple observations.

REFERENCES

1. J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal of Selected Topics in Signal Processing* **2**(5), pp. 718–726, 2008.
2. R. White, M. Postman, and M. Lattanzi, "Compression of the Guide Star digitised Schmidt plates," in *Digitized Optical Sky Surveys*, H. MacGillivray and E. Thompson, eds., pp. 167–175, Kluwer, 1992.
3. J.-L. Starck, F. Murtagh, B. Pirenne, and M. Albrecht, "Astronomical image compression based on noise suppression," *Publications of the Astronomical Society of the Pacific* **108**, pp. 446–455, 1996.
4. E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory* **52**(2), pp. 489–509, 2006.
5. D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory* **52**(4), pp. 1289–1306, 2006.
6. E. Candès, "Compressive sampling," *International Congress of Mathematics, Madrid*, 2006.
7. R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, July 2007.
8. D. Donoho and Y. Tsaig, "Extensions of compressed sensing.," *Signal Processing* **86**(3), pp. 5433–5480, 2006.
9. R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmon. Anal* **10**(1), pp. 27–44, 2001.
10. E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Preprint - available at <http://www.dsp.ece.rice.edu/cs/>*, 2006.
11. D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory* **52**(4), pp. 1289–1306, April 2006.
12. E. Candès, "The restricted isometry property and its implications for compressed sensing," tech. rep., Caltech, 2008.
13. S. Becker, J. Bobin, and E. J. Candès, "Nesta: a fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Science - submitted*, 2009.
14. A. Poglitscha, C. Waelkensb, O. Bauera, J. Cepac, H. Feuchtgrubera, T. Henning, C. van Hoofe, F. Kerschbaumf, D. Lemked, E. Renotteg, L. Rodriguez, P. Saraceno, and B. Vandenbussche, "The photodetector array camera and spectrometer (pacs) for the herschel space observatory," in *SPIE*, 2006.
15. A. N. Belbachir, H. Bischof, R. Ottensamer, F. Kerschbaum, and C. Reimers, "On-board data processing to lower bandwidth requirements on an infrared astronomy satellite: Case of herschel-pacs camera," *EURASIP Journal for Applied Signal Processing* **15**, pp. 2585–2594, 2005.

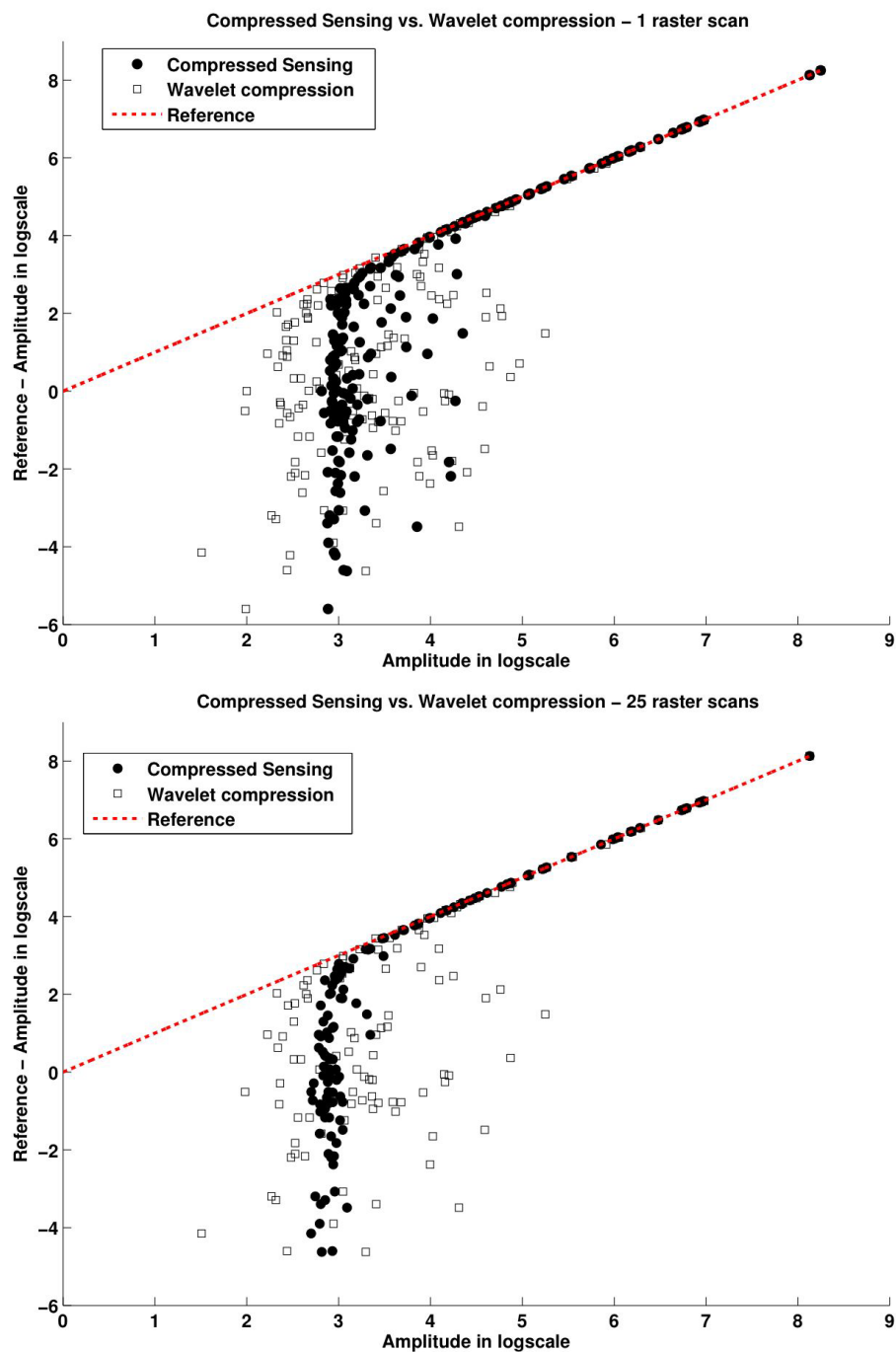


Figure 2. Photometry - Amplitude of the reference flux versus the estimated flux : comparison of Wavelet compression and CS for varying number of raster scans.

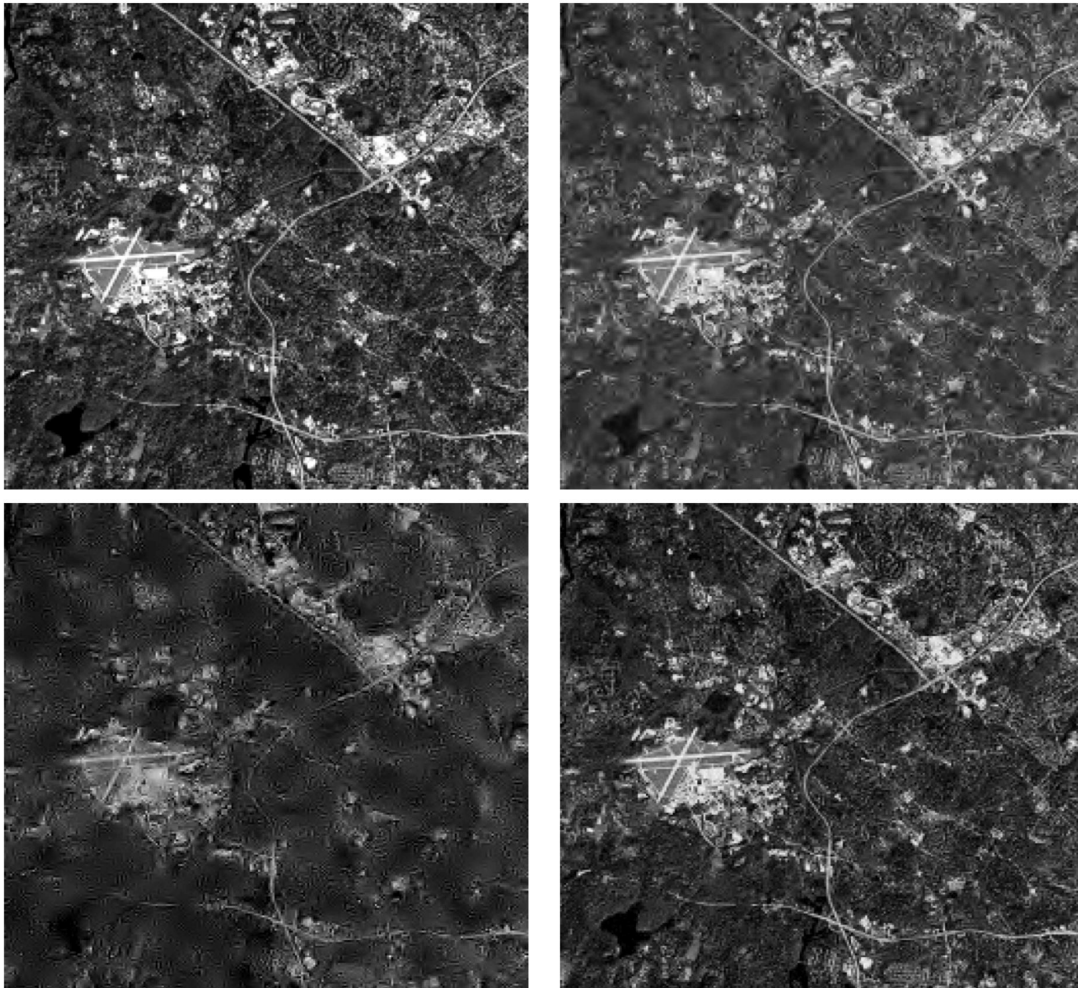


Figure 3. CS vs. Wavelet compression - First row, left picture : original 256×256 aerial images. First row, right picture : recovery after Wavelet compression with a factor 5. Second row : CS recovery from 2 and 10 observations (from left to right).

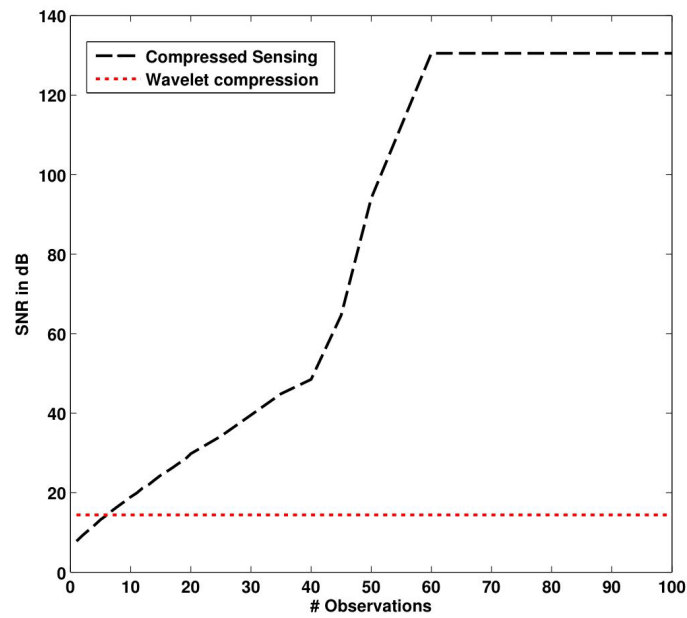


Figure 4. CS vs. Wavelet compression : Evolution of the recovery SNR for varying number of observations with a fixed compression ratio $\rho = 5$.

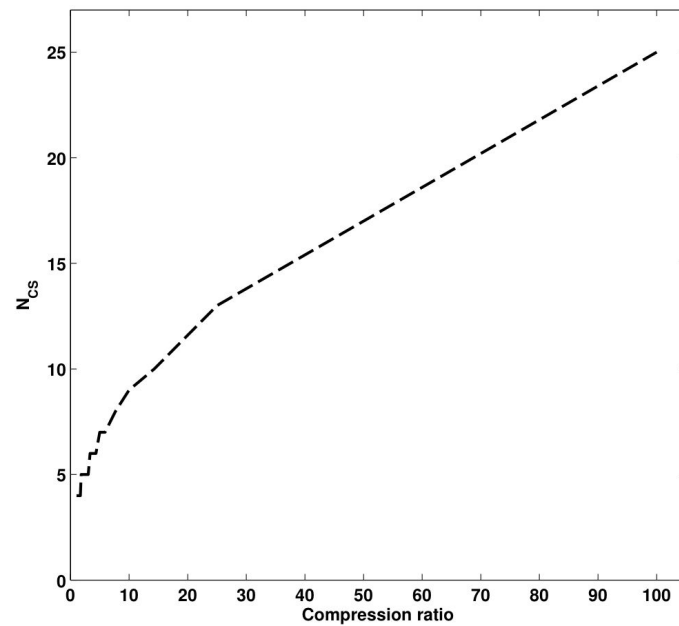


Figure 5. CS vs. Wavelet compression : Evolution of the N_{CS} , number of observations needed for CS to outperform Wavelet compression, for varying compression ratios.